Optimization of online direct marketing efforts

Xavier Drèze
André Bonfrer

Wharton · Singapore Management University · SiteSystems

Definitions

- Lucid
  - Easily understood; intelligible.
  - Mentally sound; sane or rational.
  - Translucent or transparent.
- Limpid
  - Characterized by transparent clearness; pellucid.
  - Easily intelligible; clear: *writes in a limpid style.*
  - Calm and untroubled; serene.

Test 1: Two Email campaigns

- Target:
  - Email 1: Males who had registered to sweepstakes 1.
  - Email2: Males and Females who had registered to the sweepstakes 2.
- Content:
  - Email 1: Sex and Mayhem
  - Email 2: Power and relationship

Raw Results

	Impression		Click | Impression	
	Email 1	Email 2	Email 1	Email 2
Male	42%	41%	20%	14%
Female		28%		18%

Test 2

- Target:
  - Registrants from other Video Releases.
  - Registrants from the Movie website.
- Content:
  - Video 1
  - Video 2

Test 2: Raw Results

	Video 1	Video 2
Men	20%*	14%
Women	11%	18%

* Click-through rates on targeted emails

Traditional DM Testing

- Guess a response rate
- Set a confidence level
- Set an acceptable margin of error
- Use *probability tables* to set test size
- Send test
- Wait, wait wait
- Use *probability tables* to evaluate test

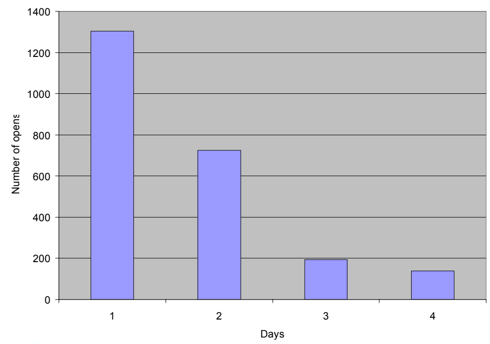
Online DM testing

- Do the same
- Do better
  - Constrained testing
  - Dynamic testing

Constrained testing

- Weekly newsletters, limited time for:
  - Writing
  - Testing
  - Sending

Open time distribution



A model of constrained testing

- M emails to send
- T periods to send the whole campaign
- r emails per hour
- 2 possible emails with click probabilities of p1, p2

- How many emails should be used for testing purposes, how long should the test last?

How long should the test last?

- If we use N emails for testing purposes then we can use T_1 periods for testing:

$$T_1 = T - \frac{M - N}{r}$$

Test size?

- How many opens (S) can we expect given N and T_1?
- What is the power of a test with size S?
- How do we balance return from testing and production?

Expected opens: S

- Simple case:
  - Send is instantaneous
  - Open time is Expo(λ)

$$f(x) = \lambda e^{-\lambda x}$$

$$F(x) = 1 - e^{-\lambda x}$$

$$\Rightarrow S = E[open \mid N, x] = N\left(1 - e^{-\lambda x}\right)$$

Send is not instantaneous

- Send rate is r, for the i^th email we have:

$$f(x,r,i) = \lambda e^{-\lambda(x-\frac{i}{r})} I_{(x,i/r)}$$

$$I_{(x,i/r)} = 1, \forall x \geq i/r$$

$$I_{(x,i/r)} = 0, \forall x < i/r$$

$$F(x,r,i) = 1 - e^{-\lambda(x-\frac{i}{r})}, if\ x > i/r$$

Expected open: S

$$S = E[open \mid N, x, r] = \sum_{i=1}^{N} F(x,r,i)$$

$$= \sum_{i=1}^{N} 1 - e^{-\lambda(x-i/r)}$$

$$= \frac{e^{\lambda/r} - e^{\lambda/r+\lambda N/r} - Ne^{\lambda x} + Ne^{\lambda/r+\lambda x}}{e^{\lambda x}(e^{\lambda/r} - 1)}$$

$$= N - \frac{(e^{\frac{\lambda N}{r}} - 1)}{e^{\lambda(x-\frac{1}{r})}\left(e^{\frac{\lambda}{r}} - 1\right)}$$

Power of test of size S

$$\alpha = \phi\left(\frac{p_1 - p_2}{\sqrt{P(1-P)\dfrac{4}{S}}}\right)$$

- where P is the average of p_1, p_2 and Φ is the CDF of the standard normal distribution.

Expected revenue for the test

$$ER(M,N,T,r,\lambda,p_1,p_2) = \underbrace{NP}_{\text{Testing}} + \underbrace{(M-N)\left(\alpha p_1 + (1-\alpha)p_2\right)}_{\text{Production}}$$

- We can then integrate over the distribution of p_1 and p_2:

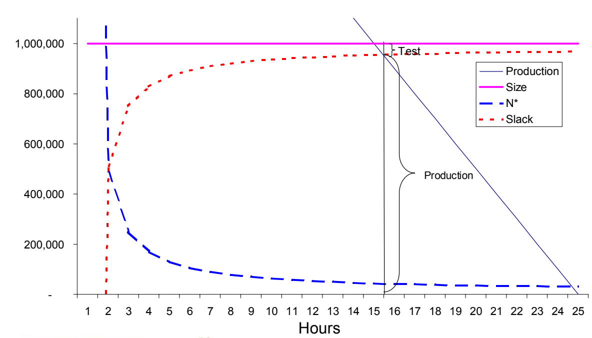
$$ER(M,N,T,r,\lambda) = \int_{P_1}\int_{P_2}\left[(M-N)(\alpha p_1 + (1-\alpha)p_2) + NP\right]P(p_1 = P_1)P(p_2 = P_2)dp_1dp_2$$

Optimal Size: N^*

- All we need to do then is to find N^* such that:

$$\max_N ER(M,N,T,r,\lambda)$$
$$s.t.: M \geq N \geq 0$$
$$M \leq Tr$$

Optimal test: Numerical Solution



Dynamic testing

- Now that we have a solution to the constrained testing situation, we can implement a dynamic one.
- As the sending occurs, one can update priors on p_1, p_2, and λ.
- As p_1, p_2, and λ are updated, one can update N and T_1.

Wharton University of Pennsylvania · Singapore Management University · SiteSystems · 20th Century Fox

Implementation issues (1)

- We have ignored the time it takes for people to read the email and click on the links.
- $T_o >> T_c$
- When solving the constrained problem, we can ignore T_c because T is in hours and T_c in tens of seconds, but if we update as we go, seconds matter!

Wharton University of Pennsylvania · Singapore Management University · SiteSystems · 20th Century Fox

Click Time: T_c

- T_c is Expo(γ)
- Ignore send time
- We can model the time to open and click as BOXMOD:

$$E[Clicks\ at\ t \mid \lambda, \gamma] = N.\frac{1}{\lambda - \gamma}[\lambda - \gamma + \gamma e^{-\lambda t} - \lambda e^{-\gamma t}]$$

Wharton University of Pennsylvania · Singapore Management University · SiteSystems · 20th Century Fox

Expected numbers of click

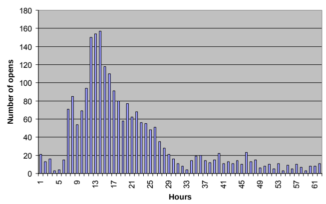
- P(open)
- P(click)

$$E[Click \mid \lambda, \gamma] = N.\frac{1}{\lambda - \gamma}[\lambda - \gamma + \gamma e^{-\lambda t} - \lambda e^{-\gamma t}].P(Open).P(Click)$$

Wharton University of Pennsylvania · Singapore Management University · SiteSystems · 20th Century Fox

Implementation issues (2)

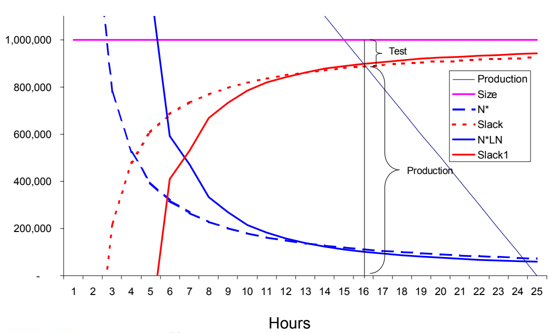
- The open times and click times look exponential from afar, but not from up close:



Time to open/click is Log-Normal

- In the short run, log-normal and expo are very different.
- There is no closed-form expression of the CDF for the log-normal.
- We have to solve numerically or maybe use a log-logistic approximation.

Log-Normal Solution



Still to do

- Solve case for log-logistic
- Solve updating formulas for P(open), P(Click), λ, and γ
- Simulate test using past data
- Simulate live test
- Live test

Enhancement

- Stopping rule for failed tests
- …